Recessive mutations in PCBD1 cause a new type of early-onset diabetes

Simaite, D ; Kofent, J ; Gong, M ; Ruschendorf, F ; Jia, S ; Arn, P ; Bentler, K ; Ellaway, C ; Kuhnen, P ; Hoffmann, G F ; Blau, N ; Spagnoli, F M ; Hubner, N ; Raile, K

Abstract: Mutations in several genes cause nonautoimmune diabetes, but numerous patients still have unclear genetic defects, hampering our understanding of the development of the disease and preventing pathogenesis-oriented treatment. We used whole-genome sequencing with linkage analysis to study a consanguineous family with early-onset antibody-negative diabetes and identified a novel deletion in PCBD1 (pterin-4 -carbinolamine dehydratase/dimerization cofactor of hepatocyte nuclear factor 1), a gene that was recently proposed as a likely cause of diabetes. A subsequent reevaluation of patients with mild neonatal hyperphenylalaninemia due to mutations in PCBD1 from the BIODEF database identified three additional patients who had developed HNF1A-like diabetes in puberty, indicating early -cell failure. We found that Pcbd1 is expressed in the developing pancreas of both mouse and Xenopus embryos from early specification onward showing colocalization with insulin. Importantly, a morpholino-mediated knockdown in Xenopus revealed that pcbd1 activity is required for the proper establishment of early pancreatic fate within the endoderm. We provide the first genetic evidence that PCBD1 mutations can cause early-onset nonautoimmune diabetes with features similar to dominantly inherited HNF1A-diabetes. This condition responds to and can be treated with oral drugs instead of insulin, which is important clinical information for these patients. Finally, patients at risk can be detected through a newborn screening for phenylketonuria.

DOI: https://doi.org/10.2337/db13-1784

Recessive mutations in *PCBD1* cause a new type of early-onset diabetes

Short title: *PCBD1* mutations cause diabetes

Deimante Simaite,[1,2] Julia Kofent,[3] Maolian Gong,[1,2] Franz Rüschendorf,[2] Shiqi Jia,[4] Pamela Arn,[5] Kristi Bentler,[6] Carolyn Ellaway,[7] Peter Kühnen,[8] Georg F. Hoffmann,[9] Nenad Blau,[9,10] Francesca M. Spagnoli,[3] Norbert Hübner,[2,11] Klemens Raile[1,11]*

[1] Experimental and Clinical Research Center (ECRC), an institutional cooperation between Charité Medical Faculty and MDC, Berlin, Berlin 13125, Germany

[2] Cardiovascular and Metabolic Diseases, Max-Delbrück Center for Molecular Medicine, Berlin 13125, Germany

[3] Laboratory of Molecular and Cellular Basis of Embryonic Development, Max-Delbrück Center for Molecular Medicine, Berlin 13125, Germany

[4] Department of Neuroscience, Max-Delbrück-Centrum for Molecular Medicine, Berlin 13125, Germany

[5] Division of Genetics, Department of Pediatrics, Nemours Children's Clinic, Jacksonville FL 32207, USA

[6] Division of Pediatric Genetics and Metabolism, Department of Pediatrics, University of Minnesota Amplatz Children's Hospital, Minneapolis MN 55454, USA

[7] Western Sydney Genetics Program, Royal Alexandra Hospital for Children, Westmead NSW 2145, Australia

[8] Institute for Experimental Pediatric Endocrinology, Charité, 13353 Berlin, Germany

Diabetes Publish Ahead of Print, published online May 21, 2014

[9] Division of Inborn Metabolic Diseases, Department of General Pediatrics, University Children's Hospital, 69120 Heidelberg, Germany

[10] Divisions of Metabolism, University Children's Hospital, 8032 Zürich, Switzerland

[11] These authors contributed equally as senior authors

*Correspondence to:

Klemens Raile MD

Experimental and Clinical Research Center (ECRC)

Charité Campus Berlin-Buch (CBB)

Lindenberger Weg 80

13125 Berlin, Germany

Telephone: +49-30-450-540-306

Fax: +49-30-450-540-926

E-mail: klemens.raile@charite.de

Word count: 2272

Number of tables: 1

Number of figures: 3

Abstract

Mutations in several genes cause non-autoimmune diabetes, but numerous patients still have unclear genetic defects, hampering our understanding of the development of the disease and preventing pathogenesis-oriented treatment. We used whole-genome sequencing with linkage analysis to study a consanguineous family with early-onset antibody-negative diabetes and identified a novel deletion in *PCBD1* (pterin-4 alpha-carbinolamine dehydratase/dimerization cofactor of hepatocyte nuclear factor-1 alpha), a gene that was recently proposed as a likely cause of diabetes. A subsequent reevaluation of patients with mild neonatal hyperphenylalaninemia due to mutations in *PCBD1* from the BIODEF database identified three additional patients who had developed *HNF1A*-like diabetes in puberty, indicating early β-cell failure. We found that Pcbd1 is expressed in the developing pancreas of both mouse and *Xenopus* embryos from early specification onward showing colocalization with insulin. Importantly, a morpholino-mediated knockdown in *Xenopus* revealed that pcbd1 activity is required for the proper establishment of early pancreatic fate within the endoderm. We provide the first genetic evidence that *PCBD1* mutations can cause early-onset non-autoimmune diabetes with features similar to dominantly inherited *HNF1A*-diabetes. This condition responds to and can be treated with oral drugs instead of insulin, which is important clinical information for these patients. Finally, patients at risk can be detected through a newborn screening for phenylketonuria.

Main Text

Diabetes is classified into type 1 diabetes (T1D) caused by autoimmune β-cell destruction, type 2 diabetes (T2D), caused by relative insulin deficiency in face of insulin resistance, gestational diabetes and other specific types, including monogenic diabetes (1). T2D is a polygenic disease with over 60 susceptibility loci and numerous risk variants co-located with genes, causing monogenic diabetes (2). Monogenic diabetes is defined by neonatal, childhood or a post-pubertal age of onset, variable clinical presentation, a lack of autoimmunity and acanthosis nigricans as well as uncommon obesity and ketoacidosis outside the neonatal period. It affects ~1% of diabetes patients (1, 3). Intrestingly, monogenic cases are often accompanied by endogenous insulin production and lack of insulin resistance. Heterozygous *GCK*, *HNF1A* and *HNF4A* mutations are the most common cause, although more than twenty other genes have been described (3, 4). However, the disease-causing variants in numerous families remain obscure. Identifying novel genes would provide insights into pathogenesis and suggest new treatment strategies for rare monogenic diabetes as well as common polygenic T1D and T2D.

We sought to identify additional genes by combining linkage analysis with the whole-genome sequencing of a consanguineous family 1 (Fig. 1A and Table 1) with non-autoimmune diabetes and no pathogenic mutations in *HNF1A*, *HNF1B, HNF4A, INS*, *ABCC8* or *KCNJ11* genes. Our institutional review boards approved the studies and written consent was obtained from the participants or their guardians. To pinpoint suggestive linkage regions, we performed a haplotype mapping and parametric linkage analysis of seven members of family 1 (Fig. 1A, red borders) using HumanCytoSNP-12 v2.1 Chip (Illumina), MERLIN software, assuming recessive inheritance, complete penetrance, and a disease allele frequency of 0.001. We obtained 24 genomic regions with positive logarithm of odds (LOD) scores, including 13 regions on 10 chromosomes with maximal LOD score of 1.3 (Supplementary Fig. 1). We next performed a whole-genome sequencing of five individuals from family 1 (Fig. 1A, asterisks) on the Complete Genomics platform (Mountain View, California, USA). The Complete Genomics pipeline was used for read mapping and allele calling.

On average, both alleles were called for ~96% genomic and ~98% exonic positions, whereas ~97% and ~99% of the called ones were covered by at least 10 reads, respectively (Supplementary Table 1). We found ~4 million small variations (small indels and SNPs) per individual genome including more than 23,000 variants in each exome.

To distinguish relevant single nucleotide polymorphisms (SNPs) and small indels from other variations, we used CGA Tools, ANNOVAR and custom scripts. We first removed intergenic variants and anticipated recessive inheritance (Supplementary Fig.2). Therefore, homozygous variants were required to be present in index case III-2, heterozygous in II-1, II-2 and I-4, but not in I-3, who is not related to I-4. Since monogenic diabetes is uncommon, we predicted the disease-causing variant to be rare and likely not yet identified. Thus, we removed all the variants present in dbSNP137, 1000 Genomes, NHLBI Exome Sequencing Project, 69 sequenced individuals from the Complete Genomics and our in-house database of non-diabetic individuals. We focused on non-synonymous variants, splice site mutations, and small indels within protein-coding regions. We further selected conserved alterations, as defined by PhastCons, and variants predicted to be deleterious by at least two tools: SIFT, Polyphen-2 or MutationTaster. This reasoning left 6 genes (Supplementary Table 4), from which we selected *PCBD1*, encoding pterin-4 alpha-carbinolamine dehydratase also known as dimerization cofactor of hepatocyte nuclear factor 1 alpha (*DCoH*) (5) for further analysis. Our decision was based on the strong expression of *PCBD1* in pancreatic islets (T1DBase and (6)), mouse pancreatic progenitors (7), and its interaction with HNF1A and HNF1B transcription factors, essential for proper pancreatic β-cell function (5, 8, 9). Moreover, *PCBD1* lay within the suggestive linkage region (Fig. 1C) and was the only gene containing a frameshift deletion. Subsequently, we validated the deletion in all diabetic members of family 1 by Sanger sequencing (Fig. 1B), using exon-intron spanning primers (Supplementary Table 2).

PCBD1 is a bifunctional protein that acts as an enzyme in the regeneration of cofactor tetrahydrobiopterin (BH_4) (10), crucial for the function of aromatic amino acid hydroxylases, and as a dimerization cofactor of transcription factors HNF1A and HNF1B (5), important in liver and

pancreas development and function. The enzymatic function of PCBD1 is defective in newborns with mild transient hyperphenylalaninemia (HPA) and high urinary levels of primapterin caused by recessive mutations (11, 12). Recently, a *PCBD1* defect was suggested to cause hypomagnesemia and diabetes (13). The novel homozygous deletion c.46del in family 1 results in a premature stop codon p.[(Lys16Cys*5)];[(Lys16Cys*5)] that abolishes the transcription factor-binding and enzymatic functions of PCBD1 (Fig. 1G).

Since III-2 developed diabetes in puberty, we reevaluated patients with neonatal hyperphenylalaninemia caused by PCBD1 who appeared in the BIODEF database (14). Three out of 7 children from 6 families, exhibiting 3 different biallelic *PCBD1* defects p.[(Glu87*)];[(Glu87*)], p.[(Glu97Lys)];[(Gln98*)], and p.[(Gln98*)];[(Gln98*)], had already developed antibody-negative diabetes with normal pancreatic morphology (Fig. 1, D-F and Table 1). Taken together, our analyses of seven independent families provide strong genetic evidence that mutations in *PCBD1* cause puberty-onset diabetes.

We tested the in vitro effects of a transient *Pcbd1* knockdown in glucose-sensitive mouse insulinoma cells using siRNA and found no obvious defects in insulin production or glucose-stimulated insulin secretion after an 80% inactivation of Pcbd1 (data not shown). Endocrine pancreas dysfunction may arise not only from β-cell inability to produce and/or secrete insulin, but also due to impaired β-cell development, proliferation and adaptation during fetal, neonatal, and pubertal age. To address whether Pcbd1 controls early pancreas development and/or pancreatic β-cell fate specification, we first examined its expression pattern in the developing pancreas of both mouse and *Xenopus* embryos. In the mouse embryo, *Pcbd1* transcript has been reported in the foregut endoderm, which contains liver and pancreas progenitors at embryonic stage (E) 8.5, and in E10.5 liver and ventral and dorsal pancreatic buds (7). We found an abundant expression of Pcbd1 in the embryonic pancreas at E12.5 and E14.5 (Fig. 2, A-D). Interestingly, Pcbd1 accumulated in endocrine progenitors, that had started to delaminate from E-cadherin-positive pancreatic epithelium and expressed insulin (Fig. 2, A' and C'). At E14.5, Pcbd1 expression was maintained in

endocrine progenitors and visible throughout the pancreatic epithelium (Fig. 2D). To analyze whether this *pcbd1* expression pattern was conserved in *Xenopus* endoderm, we performed RT-qPCR on microdissected endoderm cells. Fate map experiments in *Xenopus* have previously shown that pancreatic progenitors arise from anterior endoderm (AE), while posterior (PE) endoderm forms mainly intestine (15). We found that *pcbd1* along with the pancreatic genes *pdx1* and *ptf1a* mark future pancreatic endoderm, being expressed at higher levels in AE than in PE from the gastrula stage onward (Fig. 2E and data not shown). The pcbd1 binding partners, *hnf1a* and *hnf1b* displayed similar expression profiles, whereas the close homolog *pcbd2* was almost absent from AE cells (Fig. 2E). Furthermore, an *in situ* hybridization in *Xenopus* embryos showed *pcbd1* expression in pancreatic rudiments, overlapping with the expression pattern of *insulin* (Fig. 2, F-I). Overall, these results indicate that Pcbd1 is expressed in the developing pancreas of both mouse and *Xenopus* embryos, suggesting a potential evolutionarily conserved function.

Pcbd1 knockout mouse showed mild glucose intolerance (16), although no pancreas-specific function had yet been assigned to Pcbd1. To determine if Pcbd1 influences early pancreas fate specification, we undertook a loss-of-function approach in *Xenopus*. A specific morpholino oligonucleotide (pcbd1-MO) was designed to block *pcbd1* pre-mRNA splicing. The injection of pcbd1-MO into AE cells of the eight-cell stage *Xenopus* embryos resulted in a dose-dependent downregulation of *pcbd1* mRNA (Fig. 3A and data not shown), accompanied by a significant reduction in the expression of pancreatic progenitor genes *pdx1*, *ptf1a, sox9,* as well as *insulin* (Fig. 3A). Moreover, *hnf1a* and *hnf1b* and the hnf1 target gene *fibrinogen* were downregulated upon pcbd1-MO injection, whereas *pcbd2* mRNA levels remained unchanged (data not shown). *In situ* hybridization of pcbd1-MO injected *Xenopus* embryos corroborated these observations, showing a strong reduction or complete loss of *ptf1a* expression in both dorsal and ventral pancreatic buds, but not in the eye and hindbrain (Fig. 3B-D). Taken together, these results indicate that pcbd1 activity within the endoderm is required for the proper establishment of the pancreatic region in vertebrates.

We examined 8 patients with inherited biallelic *PCBD1* mutations from 7 families: one case by whole-genome sequencing combined with linkage analysis, which identified a novel deletion in the *PCBD1* gene and seven cases by recalling the patients from the BIODEF database (14). Interestingly, two diabetic patients (family 1, III-2 and family 3, II-2) had normal phenylalanine levels, suggesting that the enzymatic function of PCBD1 had been compensated for by PCBD2, as has been proposed for mouse (16).

Insulin was the first option of treatment of *PCBD1*-diabetes, subsequently replaced by sulphonylureas or glinides. This brisk response to oral drugs resembles the patients with *HNF1A*-diabetes (17). *PCBD1*-diabetes manifests earliest in puberty as no younger cases that were investigated were diabetic (Table 1). Remarkably, mid-adulthood-onset T2D developed in 4 out of 7 families (families 1, 3, 4 and 6; Table 1 and Supplementary Table 5). We confirmed heterozygous *PCBD1* defects in affected individuals of family 1 and 3. Therefore, it is likely that monoallelic variants in *PCBD1* increase the risk of T2D, as has been shown for other monogenic diabetes genes (2). Moreover, it is suggestive that *PCBD1* mutations increase T2D susceptibility specifically when combined with other risk factors such as excess weight and age. This is implied by the fact that only overweight/obese heterozygotes developed T2D, whereas none of those with a normal BMI did (Supplementary Table 5). Furthermore, the two overweight/obese parents who did not develop diabetes were 32 and 34 years old and thus had not yet reached the age of the onset of diabetes in the other heterozygotes.

We investigated two major mechanisms of diabetes: the production and secretion of insulin in insulinoma cells and the early regulation of pancreatic and β-cell specification in vertebrates. As PCBD1 has been described to enhance HNF1A activity on some promoters (5), we examined whether a transient inactivation of PCBD1 decreases insulin production and secretion in vitro, but failed to see any relevant effect (data not shown). Previous studies have shown that *Pcbd1* is abundantly expressed throughout embryonic development in both *Xenopus* and mammalian embryos (18, 19). Interestingly, we found that PCBD1 is expressed in mouse pancreatic progenitors

(7) and delaminating endocrine cells from very early stages onward (Fig. 2, A-D). This spatiotemporal expression of pcbd1 is also conserved in *Xenopus* embryos (Fig. 2, G and I), suggesting conserved regulatory functions. In line with this, *Xenopus pcbd1* morphants exhibited a reduced expression of endodermal and pancreatic transcription factors, indicating defects in early pancreas specification. Notably, both dorsal and ventral pancreatic rudiments fail to be established in pcbd1-depleted embryos, as judged by the absence of *Ptf1a* expression (Fig. 3, B-D). Altogether, these findings suggest an early role of pcbd1 in establishing the pancreatic progenitor pool during embryogenesis, which might lead to a reduced pancreatic β-cell mass in the adult. The human *PCBD1*-diabetes phenotype with developing insulin-deficiency in the face of somatic growth and weight gain is in line with the notion that an intrinsic program established early in development, is critical in determining the final size of the pancreas and is not subject to growth compensation (20).

Transcription factor HNF1B regulates early pancreatic development in mouse and human and is stabilized by PCBD1 (5, 21). A lack of PCBD1 might impair the HNF1B-mediated establishment of pancreatic cell fate during embryogenesis. In *Xenopus* embryos, this hypothesis is supported by a reduced expression of *hnf1b* and its direct target genes, such as *fibrinogen,* upon pcbd1-MO injection. However, a *Pcbd1* knockout mouse shows a relatively mild phenotype compared to *Xenopus pcbd1* morphant or human *HNF1A* and *HNF1B* loss-of-function phenotypes (8, 9). These differences might be due to a partial functional redundancy between Pcbd1 and the close homologue Pcbd2 in the mouse (16), which does not seem sufficient to prevent diabetes. In line with this, pcbd2 is expressed at low levels in the endoderm and is almost undetectable in the pancreatic territory of *Xenopus* embryos (Fig. 2E). Further experiments are required to dissect the mechanisms underlying the Pcbd1 loss of function phenotype and Pcbd2's role in its development.

Recent observations have suggested that a gradual loss of transcription factors Pdx1, Nkx6.1 and MafA, crucial for early pancreas development, leads to a destabilized adult β-cell state and an exhaustion of function, possibly contributing to the pathogenesis of T2D (22). The HNF1 family of transcription factors also controls both aspects. HNF1B is required to set up the early pancreatic

transcriptional program (21) and HNF1A maintains a proper transcriptional network in mature β-cells (23). Thus, their binding partner PCBD1 might have different effects on the pancreas at different time points. Finally, PCBD1 mutations probably cause hyperphenylalaninemia and diabetes by affecting not one particular pathway but several genetic, metabolic and signaling programs in different tissues, as shown for *HNF1A* gene (24). Future studies will aim to fully understand how PCBD1 regulates β-cell functions and the mechanisms leading to diabetes.

In summary, we provide the first genetic evidence that *PCBD1* mutations can cause early-onset monogenic diabetes. We recommend a monitoring of neonatal hyperphenylalaninemia patients and their relatives with *PCBD1* mutations in puberty and later in life for an occurrence of diabetes. We suggest screening *HNF1A*-like diabetes cases without mutations in *HNF1A* and *HNF4A* genes for recessive alterations in *PCBD1*, since they can be treated with oral antidiabetic drugs. Larger numbers of *PCBD1*-diabetes patients will be needed to determine how commonly insulin can be replaced with sulfonylureas, as larger studies of individuals with *HNF1A*, *KCNJ11* and *ABCC8* defects show that this treatment is successful in most but not all the cases.

Acknowledgements

We thank our patients and their families. We also thank Nadine Wittstruck, Sabine Schmidt and Mathias Gerhard (Cardiovascular and Metabolic Diseases, Max-Delbrück Center for Molecular Medicine, Berlin 13125, Germany) for superb technical support, Dr. Patrick Ferreira (Division of Medical Genetics, Alberta Children's Hospital, Calgary AB T3B 6A8, Canada), Dr. Roger Germann (Children's Hospital, Klinikum Karlsruhe, Karlsruhe 76042, Germany) and Dr. Johannis B.C. de Klerk (Department of Pediatrics, Sophia Children's Hospital, Erasmus Medical Center, Rotterdam CB 3000, the Netherlands) for supporting the BIODEF recall and providing additional family information as well as Prof. Friedrich C. Luft (Experimental and Clinical Research Center, Berlin 13125, Germany) and Russel Hodge (Max-Delbrück Center for Molecular Medicine, Berlin 13125, Germany) for reviewing the manuscript. The ECRC supports Klemens Raile. The Deutsche Forschungsgemeinschaft supported Maolian Gong (DFG: GO 1990/1-1). FM Spagnoli is supported by the Helmholtz Association (VH-NG-425); ERC-2009-St. Grant HEPATOPANCREATIC (243045); and FP7 IRG-ENDOPANC (239534). No potential conflicts of interest relevant to this article were reported.

D.S. designed, planned and carried out experiments, analyzed the data and wrote the manuscript, J.K. planned and carried out mouse and *Xenopus* experiments, analyzed the data, M.G. planned experiments, discussed the data and reviewed the manuscript, F.R. performed linkage analysis and discussed the data, S.J. performed mouse cell culture experiments, P.A., K.B., C.E., P.K. and G.K.H. provided clinical data and contributed to the discussion, N.B. planned and carried out the recall of patients and their families from the BIODEF database, provided clinical data, reviewed and edited the manuscript, F.M.S. designed mouse and *Xenopus* experiments, discussed the data, reviewed and edited the manuscript, N.H. designed the study, discussed the data, reviewed and edited the manuscript, K.R. designed the study, followed-up patients, performed clinical examinations and discussed the data, reviewed and edited the manuscript. K.R is the guarantor of

this work, and as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

References

1. American Diabetes A. Diagnosis and classification of diabetes mellitus. Diabetes Care 2013;36 Suppl 1:S67-74
2. Morris AP, Voight BF, Teslovich TM, et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nat Genet 2012;44:981-990
3. Ashcroft FM, Rorsman P. Diabetes mellitus and the beta cell: the last ten years. Cell 2012;148:1160-1171
4. Ellard S, Lango Allen H, De Franco E, et al. Improved genetic testing for monogenic diabetes using targeted next-generation sequencing. Diabetologia 2013;56:1958-1963
5. Mendel DB, Khavari PA, Conley PB, et al. Characterization of a cofactor that regulates dimerization of a mammalian homeodomain protein. Science 1991;254:1762-1767
6. Eizirik DL, Sammeth M, Bouckenooghe T, et al. The human pancreatic islet transcriptome: expression of candidate genes for type 1 diabetes and the impact of pro-inflammatory cytokines. PLoS Genet 2012;8:e1002552
7. Rodriguez-Seguel E, Mah N, Naumann H, et al. Mutually exclusive signaling signatures define the hepatic and pancreatic progenitor cell lineage divergence. Genes Dev 2013;27:1932-1946
8. Lindner TH, Njolstad PR, Horikawa Y, Bostad L, Bell GI, Sovik O. A novel syndrome of diabetes mellitus, renal dysfunction and genital malformation associated with a partial deletion of the pseudo-POU domain of hepatocyte nuclear factor-1beta. Hum Mol Genet 1999;8:2001-2008
9. Yamagata K, Oda N, Kaisaki PJ, et al. Mutations in the hepatocyte nuclear factor-1alpha gene in maturity-onset diabetes of the young (MODY3). Nature 1996;384:455-458
10. Citron BA, Davis MD, Milstien S, et al. Identity of 4a-carbinolamine dehydratase, a component of the phenylalanine hydroxylation system, and DCoH, a transregulator of homeodomain proteins. Proc Natl Acad Sci U S A 1992;89:11891-11894
11. Thony B, Neuheiser F, Kierat L, et al. Hyperphenylalaninemia with high levels of 7-biopterin is associated with mutations in the PCBD gene encoding the bifunctional protein pterin-4a-carbinolamine dehydratase and transcriptional coactivator (DCoH). Am J Hum Genet 1998;62:1302-1311
12. Thony B, Neuheiser F, Kierat L, et al. Mutations in the pterin-4alpha-carbinolamine dehydratase (PCBD) gene cause a benign form of hyperphenylalaninemia. Hum Genet 1998;103:162-167
13. Ferre S, de Baaij JH, Ferreira P, et al. Mutations in PCBD1 Cause Hypomagnesemia and Renal Magnesium Wasting. J Am Soc Nephrol 2013;
14. Opladen T, Hoffmann GF, Blau N. An international survey of patients with tetrahydrobiopterin deficiencies presenting with hyperphenylalaninaemia. J Inherit Metab Dis 2012;35:963-973
15. Chalmers AD, Slack JM. The Xenopus tadpole gut: fate maps and morphogenetic movements. Development 2000;127:381-392
16. Bayle JH, Randazzo F, Johnen G, et al. Hyperphenylalaninemia and impaired glucose tolerance in mice lacking the bifunctional DCoH gene. J Biol Chem 2002;277:28884-28891
17. Awa WL, Thon A, Raile K, et al. Genetic and clinical characteristics of patients with HNF1A gene variations from the German-Austrian DPV database. Eur J Endocrinol 2011;164:513-520
18. Strandmann EP, Senkel S, Ryffel GU. The bifunctional protein DCoH/PCD, a transcription factor with a cytoplasmic enzymatic activity, is a maternal factor in the rat egg and expressed tissue specifically during embryogenesis. Int J Dev Biol 1998;42:53-59
19. Pogge yon Strandmann E, Ryffel GU. Developmental expression of the maternal protein XDCoH, the dimerization cofactor of the homeoprotein LFB1 (HNF1). Development 1995;121:1217-1226
20. Stanger BZ, Tanaka AJ, Melton DA. Organ size is limited by the number of embryonic progenitor cells in the pancreas but not the liver. Nature 2007;445:886-891
21. Barbacci E, Reber M, Ott MO, Breillat C, Huetz F, Cereghini S. Variant hepatocyte nuclear factor 1 is required for visceral endoderm specification. Development 1999;126:4795-4805
22. Guo S, Dai C, Guo M, et al. Inactivation of specific beta cell transcription factors in type 2 diabetes. J Clin Invest 2013;

23. Shih DQ, Stoffel M. Dissecting the transcriptional network of pancreatic islets during development and differentiation. Proc Natl Acad Sci U S A 2001;98:14189-14191
24. Servitja JM, Pignatelli M, Maestro MA, et al. Hnf1alpha (MODY3) controls tissue-specific transcriptional programs and exerts opposed effects on cell growth in pancreatic islets and liver. Mol Cell Biol 2009;29:2945-2959
25. Spagnoli FM, Brivanlou AH. The Gata5 target, TGIF2, defines the pancreatic region by modulating BMP signals within the endoderm. Development 2008;135:451-461

Table 1. Clinical characteristics of patients with biallelic *PCBD1* mutations investigated in present study.

Patient Number	**ECRC 7, Family 1, III-2**	**BIODEF§ 319, Family 2, II-1**	**BIODEF§ 344, Family 3, II-2**	**BIODEF§ 273, Family 3, II-1**	**BIODEF§ 272, Family 4, II-1**	**BIODEF§ 329, Family 5**	**BIODEF§ 701, Family 6**	**BIODEF§ 620, Family 7**
Sex	F	F	F	M	M	M	F	F
Present age, years	20	17	21	20	20	18	1	4
Ethnicity	Turkish	Caucasian	Ashkenazi Jewish	Ashkenazi Jewish	Caucasian	Turkish	Caucasian	Caucasian
Consanguinity	Yes	No	No	No	No	Yes	No	No
Nucleotide aberration*	c.[(46del)]; [(46del)]	c.[(289G>A)]; [(292C>T)]	c.[(259G>T)]; [(259G>T)]	c.[(259G>T)]; [(259G>T)]	c.[(292C>T)]; [(292C>T)]	c.[(79G>T;263G>A)]; [(79G>T;263G>A)]	c.[(292C>T)]; [(292C>T)]	c.[(292C>T)]; [(213_215del)]
Protein alteration	p.[(Leu16Cysfs*5)]; [(Leu16Cysfs*5)]	p.[(Glu97Lys)]; [(Gln98*)]	p.[(Glu87*)]; [(Glu87*)]	p.[(Glu87*)]; [(Glu87*)]	p.[(Gln98*)]; [(Gln98*)]	p.[(Glu27*;Asp88Gln)]; [(Glu27*;Asp88Gln)]	p.[(Gln98*)]; [(Gln98*)]	p.[(Asn71del)]; [(Gln98*)]
Variation ID†	Novel	(CM981485); rs121913015 (CM981486)	rs104894172 (CM930575)	rs104894172 (CM930575)	rs121913015 (CM981486)	CM981482; rs115117837 (CM981484)	rs121913015 (CM981486)	rs121913015 (CM981486); Novel
Diabetes	Yes	Yes	Yes	No	Yes	No	No	No
Onset, years	14	15	12	-	18	-	-	-
Symptoms	Glucosuria, polyuria	Polyuria, polydipsia	Glucosuria	-	Polyuria, polydipsia	-	-	-
IBGL, mg/dl	275	270	262	-	414	-	69	89
Initial BMI, (z-score)	23.4 (+2.1)	20.2 (+0.4)	21.3 (+1.4)	-	20.1 (-0.4)	26.3 (+0.9)	16.5 (+0.1)	17.3 (+1.1)
Initial HbA1c, % (mmol/mol)	7.8 (62)	6.5 (48)	<6.1 (43)	-	14.6 (136)	4.9 (30)	4.3 (23)	4.9 (30)
β-cell autoantibodies	Negative	Negative	Negative	N.D.	Negative	N.D.	N.D.	N.D.
Pancreas morphology‡	Normal	N.D.	N.D.	N.D.	Normal	Normal	N.D.	N.D.
Treatment	Insulin, meglitinide	Insulin, sulphonylurea	Insulin, sulphonylurea, lifestyle	-	Insulin, sulfonylurea	-	-	-
Family history	Father, mother and maternal grandmother T2D	Both parents healthy	Mother healthy, father T2D, obesity	Mother healthy, father T2D, obesity	Both parents healthy, maternal grandparents T2D	Both parents healthy	Both parents healthy, maternal grandfather T2D	Both parents healthy
HPA status	No	Yes	No	Yes	Yes	Yes	Yes	Yes
Reference	Unpublished	10, 12	11	11	11, 12	10, 12	Unpublished	Unpublished

HPA – hyperphenylalaninemia during neonatal period, BMI – body mass index (kg/m²), IBGL – initial blood glucose levels, HbAc1 – glycated hemoglobin, T2D – type 2 diabetes mellitus, N.D. – not done
*Positions refer to Consensus CDS database accession number 31217.1. RefSeq number for human *PCBD1* mRNA is NM_000281.2.
†dbSNP and/or HGMD (in brackets) database accession numbers
‡Determined by abdominal ultrasound
§BIODEF database

Figure Titles and Legends

Figure 1. Mutations in *PCBD1* cause early-onset diabetes.

(A) Pedigree of family 1 with proband III-2 having an early-onset diabetes due to homozygous mutation p.[(L16Cfs*5)];[(L16Cfs*5)]. Three of the tested relatives (I-4, II-1, II-2) have the same heterozygous mutation and diabetes. Here and in the following panels, individuals with adolescent-onset diabetes are marked in black and with mid-adulthood-onset diabetes in grey. HPA stands for hyperphenylalaninemia. Red borders mark individuals used in linkage analysis and red asterisks mark individuals whose whole genome was sequenced. **(B)** Spherograms of direct-sequenced individuals of family 1 showing the homozygous mutation in the proband III-2 and heterozygous mutation in three family members with diabetes. **(C)** Linkage analysis of family 1 showing the *PCBD1* mutation (arrow) in the chromosome 10 linkage region with expected maximal LOD score of 1.3. **(D)** Pedigree of family 2 with proband II-1 suffering with adolescent-onset diabetes and carrying a compound heterozygous mutation p.[(Gln97Lys)];[(Gln98*)]. Both parents who possess one diseased allele are healthy. **(E)** Pedigree of family 3 with both children bearing the biallelic p.[(Gln87*)];[(Gln87*)] mutation. Only II-2 developed adolescent-onset diabetes so far. Father, a heterozygous mutation carrier, developed diabetes later in life, whereas mother is healthy. **(F)** Pedigree of family 4. The proband II-1, possessing a homozygous p.[(Gln98*)];[(Gln98*)] mutation, developed diabetes early in life. His heterozygous parents are healthy. **(G)** Location of known disease-causing mutations in *PCBD1* cDNA and protein. Mutations marked in red manifested as puberty-onset diabetes. Arrow marks the mutation first identified in our study. Blue lines show the positions of His62, His63 and His80 residues present in the active center of PCBD1 enzyme.

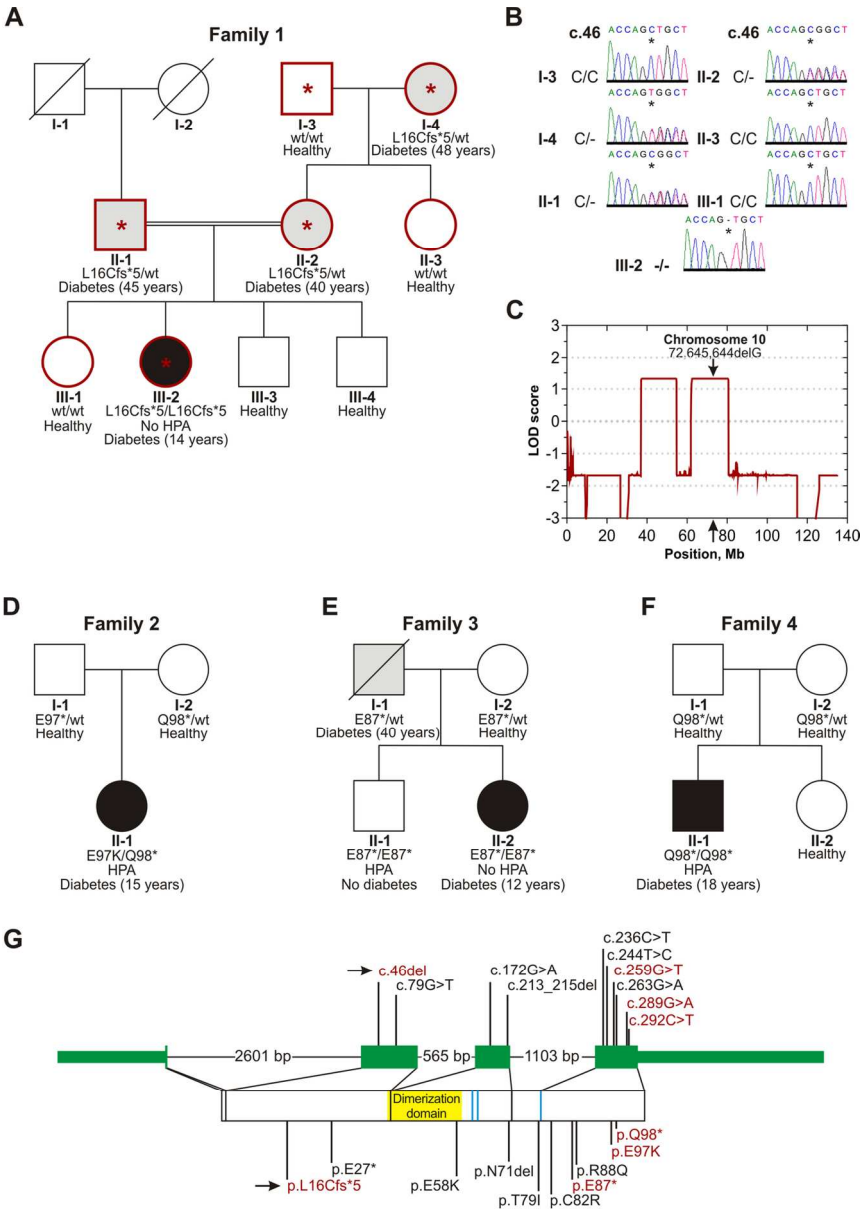
Figure 2. Pcbd1 in mouse and *Xenopus* embryonic pancreas.

(A-D) Immunofluorescence analysis of Pcbd1, Pdx1, Insulin (Ins) and E-cadherin (Ecad) on E12.5 (A, C) and E14.5 (B, D) 10 μm cryosections of mouse pancreas epithelium. Dashed boxes indicated

delaminating endocrine cells, which displayed strong nuclear and cytoplasmic Pcbd1 expression and coexpressed Ins (C' and D) or Pdx1 (A' and B). Bar, 50 μm. **(E)** Posterior endoderm (PE) and anterior endoderm (AE) explants were dissected at early gastrula stage, cultured until indicated stages and assayed for expression of the indicated genes by RT-qPCR analysis. At all stages analyzed *pcbd1* expression mirrored that of anterior endodermal markers, such as *foxa2*, *hnf1a* and *hnf1b*, and of pancreatic transcription factors, such as *pdx1* and *ptf1a*. Data were normalized to that of *ODC* and represented as fold changes compared to PE sample (set to 1). **(F-I)** Whole-mount *in situ* hybridization for *pcbd1* in *Xenopus* embryos and dissected gut (performed as described previously (25)). *Pcbd1* transcript was detected in dorsal pancreas (G, I), overlapping with insulin-expression domain (F, H). Abbreviations: dp, dorsal pancreas; lv, liver; pa, pancreas, pn, pronephros. Bar, 1 mm. Error bars represent ± SD. Asterisks stand for p-values calculated using Student's t-test: ns>0.05, *<0.05, **<0.01, ***<0.001.

Figure 3. *Pcbd1*-morpholino-knockdown in *Xenopus* anterior endoderm explants

(A) RT-qPCR analysis of pcbd1-MO-injected AE explants. Pcbd1-MO (10 ng) (GeneTools LCC, Philomath, USA) was injected into two vegetal dorsal blastomeres of 8-cell stage *Xenopus* embryos, AE explants were dissected at gastrula stage and assayed at tadpole stage for the indicated pancreatic and hepatic genes by RT-qPCR assay. Abbreviations, *insulin* (*ins*), *fibrinogen* (*fgn*). Data were normalized to that of *ODC* and represented as fold changes compared to AE uninjected control sample (set to 1). (**B-D)** Whole-mount *in situ* hybridization analysis of *ptf1a* in control and pcbd1-MO-injected *Xenopus* embryos. (B', C' and D') Vibratome transverse sections through the ventral pancreatic (vp) rudiment stained for *ptf1a*. (B'', C'' and D'') Vibratome transverse sections through the dorsal pancreatic (dp) rudiment stained for *ptf1a*. Dashed lines indicate the cross-sectional planes. Bar, 1mm. Error bars represent ± SD. Asterisks stand for p-values calculated using Student's t-test: ns>0.05, *<0.05, **<0.01, ***<0.001.



121x170mm (300 x 300 DPI)

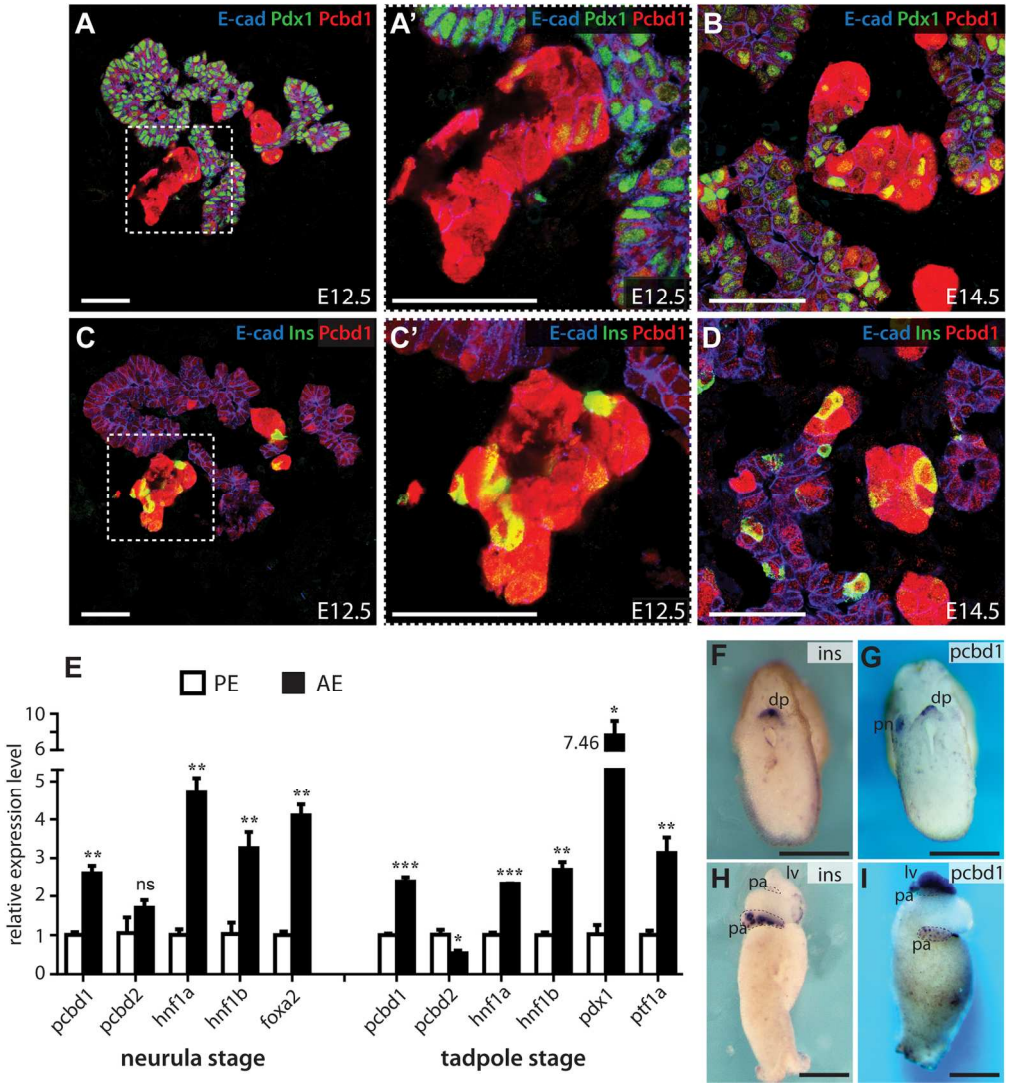


Figure 2
193x209mm (300 x 300 DPI)

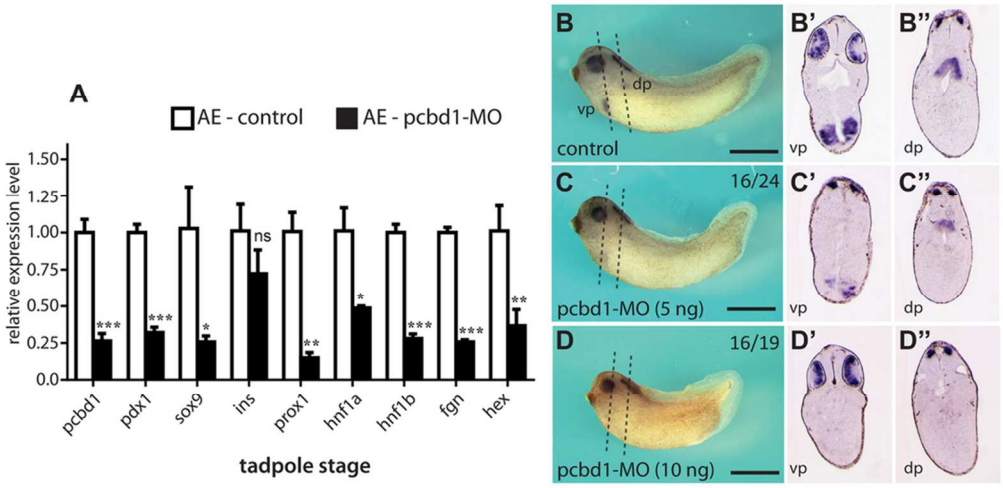


Figure 3
89x44mm (300 x 300 DPI)

Recessive mutations in *PCBD1* cause a new type of early-onset diabetes

SUPPLEMENTARY DATA

Supplementary Table 1. Statistics of whole-genome sequencing performance of family 1.

Sample	Fully called genomic positions, %	% of fully called genomic positions covered at ≥10x	Total genomic variants	Fully called exonic positions, %	% of fully called exonic positions covered at ≥10x	Total exonic variants
I-I	95.6	96.0	4,091,829	98.2	98.6	23,500
I-II	95.7	96.6	4,141,813	98.1	98.4	23,915
II-III	95.8	96.4	4,111,511	98.4	98.8	23,624
II-II	95.8	96.8	4,138,100	98.2	98.6	23,760
III-II	96.0	97.4	4,049,355	98.4	98.9	23,282

Supplementary Table 2. PCR primers used to amplify *PCBD1* exons for Sanger sequencing.

Primer name	Primer sequence	Exon	Annealing temperature
PCBD1_1bF	gatggtctcacgagggaaaca	1	60
PCBD1_1aR	gcaggggactcgaaaagact		
PCBD1_2F	cccagcctattgctcaaaga	2	58
PCBD1_2R	ctggatgagtgtggtgtctga		
PCBD1_3F	aggatgtcaaggggaaatg	3	58
PCBD1_3R	aggcatgtgcaatctcagtg		
PCBD1_4F	actggccagctgctattctg	4	58
PCBD1_4R	ttggattgacctgtggaaaag		

Supplementary Table 3. *Xenopus* primers used for SyberGreen RT-qPCR.

Gene	Forward primer	Reverse primer
odc	ttcgggtgattccttgccac	gccattgtgaagactctctccaatc
pcbd1	catgacaagggtggctctcc	ctacatcactattggtaatgtgtttcct
pcbd2	ctttaaccaggcatttggatt	gtgtagtcagagttatctgaac
hnf1a	ccatggcaaaacttatggatttaga	ggagatggggtactctgactg
hnf1b	gaagaaagagaagctttagtgg	gactatatctcagcccttgc
pdx1	gttccctcagctgcttatcg	taccaaggggttgctgtagg
ptf1a	atggaaacggtcctggagca	gaggatgagaaggagaagttg
insulin	aggcttcttctactaccctaag	acaatccccctcttcattt
sox9a	caactaattgcgcactgggg	tcttcagcaaaggcacccaa
foxa2	taccaacatcaactccatgagc	gtaacttcgccggtaagttttg
prox1	ctgatatctcaccttattcgg	tgggaggtgatgcatctgttg
hex	cctttccgcttgtgcagagg	aacagcgcatctaatgggac
fibrinogen	aagatgactcagtgggcagc	ttcaatgccgccttctcctt

Supplementary Table 4. Final set of homozygous mutations from whole-genome sequencing dataset of family 1 after selection of novel, protein affecting, conserved and damaging variants.

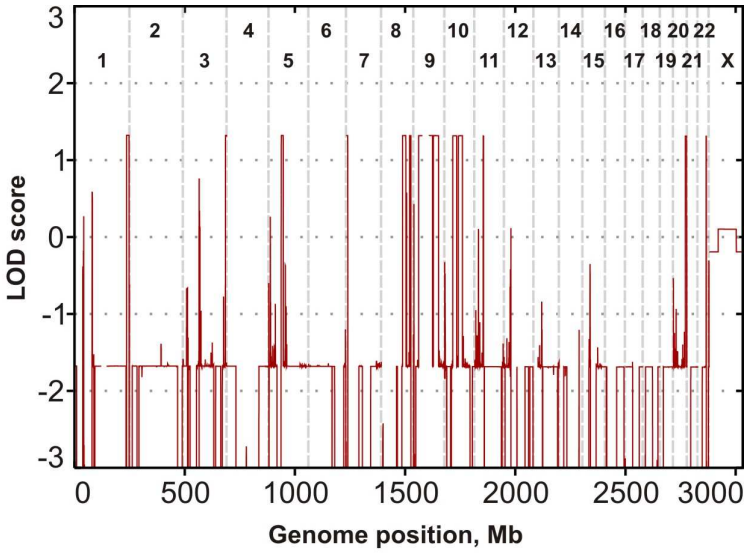
Gene	Chr	Start	End	Reference	Observed	SIFT*	PP2†	MT‡	Linkage	T1DB§	HI‖	Vpa8.5¶	Dpa8.5#	Rank
RECK	9	36091226	36091226	G	A	T	P	D	Yes	Enriched	1.5	9.4	14.2	4
GDF10	10	48426745	48426745	T	C	D	D	D	Yes	Low	0.8	2.1	0.4	5
OGDHL	10	50953532	50953532	C	T	D	D	D	Yes	Moderate	0.5, 1.2, 4.5	16.9	15.7	3
MYPN	10	69955261	69955261	C	T	D	D	D	Yes	Low	0.6	NE**	NE**	6
PCBD1	10	72645644	72645644	G	-	D	NA	D	Yes	Enriched	47.2	72.0	40.6	1
CTSZ	20	57571717	57571717	G	A	D	D	D	Yes	Moderate	152.9	64.7	96.9	2

*SIFT – SIFT prediction (T – tolerant, D - deleterious)
†PP2 – Polyphen-2 prediction (P – possibly damaging, D – probably damaging, NA – not available)
‡MT – MutationTaster prediction (D – disease-causing)
§T1DB – expression in human islets present T1DBase (array data)
‖HI – RNAseq RPKM (Reads Per Kilobase of transcript per Million mapped reads) values in human islets (5)
¶Vpa8.5 – RNAseq FPKM (Fragments Per Kilobase of transcript per Million mapped reads) values in mouse ventral pancreatic bud at E8.5 (6)
#Dpa – 8.5 RNAseq FPKM (Fragments Per Kilobase of transcript per Million mapped reads) values in mouse dorsal pancreatic bud at E8.5 (6)
**NE – not expressed

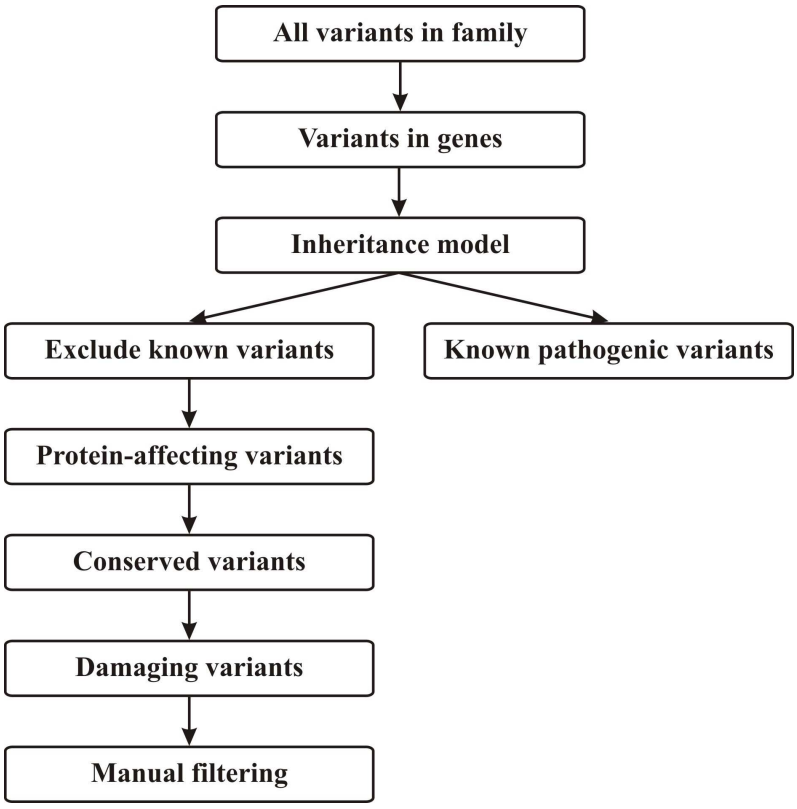
Supplementary Table 5. Characteristics of family members with heterozygous *PCBD1* mutations.

Family	Relationship	Protein alteration	Present age, years	Ethnicity	Diabetes				Weight range*(BMI, kg/m²)
					Diabetes, type	Onset, years	Recent HbA1c, % (mmol/mol)	Treatment	
1	Father	p.[(Leu16Cysfs*5)];[(wt)]	49	Turkish	Yes, T2D	45	6.2 (44.3)	Lifestyle	Obese (31.7)
	Mother	p.[(Leu16Cysfs*5)];[(wt)]	48	Turkish	Yes, T2D	40	6.9 (51.9)	Metformin	Overweight (29.6)
	Maternal grandmother	p.[(Leu16Cysfs*5)];[(wt)]	69	Turkish	Yes, T2D	48	6.9 (51.9)	Metformin, then SU	Overweight (28.5)
2	Father	p.[(Glu97Lys)];[(wt)]	47	Caucasian	No	N.A.	N.A.	N.A.	Normal (< 25)
	Mother	p.[(Gln98*)];[(wt)]	48	Caucasian	No	N.A.	N.A.	N.A.	Normal (< 25)
3	Father	p.[(Glu87*)];[(wt)]	Died at 50	Ashkenazi Jewish	Yes, T2D	35	N.A.	Metformin, then SU	Obese (>30)
	Mother	p.[(Glu87*)];[(wt)]	55	Ashkenazi Jewish	No	N.A.	N.A.	N.A.	Normal (< 25)
4	Father	p.[(Gln98*)];[(wt)]	36	Caucasian	No	N.A.	N.A.	N.A.	Normal (< 25)
	Mother	p.[(Gln98*)];[(wt)]	34	Caucasian	No	N.A.	N.A.	N.A.	Normal (< 25)
5	Father	p.[(Glu27*;Asp88Gln)];[(wt;wt)]	45	Turkish	No	N.A.	N.A.	N.A.	Normal (< 25)
	Mother	p.[(Glu27*;Asp88Gln)];[(wt;wt)]	40	Turkish	No	N.A.	N.A.	N.A.	Normal (< 25)
6	Father	p.[(Gln98*)];[(wt)]	36	Caucasian	No	N.A.	N.A.	N.A.	Normal (< 25)
	Mother	p.[(Gln98*)];[(wt)]	34	Caucasian	No	N.A.	N.A.	N.A.	Normal (< 25)
7	Father	p.[(Gln98*)];[(wt)]	34	Caucasian	No	N.A.	N.A.	N.A.	Overweight (29.1)
	Mother	p.[(Asn71del)];[(wt)]	32	Caucasian	No	N.A.	N.A.	N.A.	Obese (31.2)

BMI – body mass index (kg/m²), HbAc1 – glycated hemoglobin, T2D – type 2 diabetes mellitus, SU – sulfonylurea, N.A. – not available
*Weight range (Normal – BMI < 25 kg/m², Overweight – BMI 25-30 kg/m², Obese – BMI > 30 kg/m²)



Supplementary Figure 1. Linkage analysis of family 1. 24 genomic regions have positive LOD score, including 13 regions reaching maximal expected LOD score of 1.3, in case of perfect recombination.



Supplementary Figure 2. Filtering strategy employed to reduce all variants of family 1 down to a single candidate gene.

Web resources

MERLIN http://www.sph.umich.edu/csg/abecasis/merlin/index.html

ANNOVAR http://www.openbioinformatics.org/annovar/

Human Genome Reference http://www.ncbi.nlm.nih.gov/refseq/

CGA Tools http://cgatools.sourceforge.net

dbSNP http://www.ncbi.nlm.nih.gov/SNP

NHLBI Exome Sequencing Project (ESP) Exome Variant Server https://esp.gs.washington.edu/EVS

1000 Genomes Project www.1000genomes.org

69 sequenced individuals by Complete Genomics www.completegenomics.com/public-data/69-Genomes

PhastCons scores http://hgdownload-test.cse.ucsc.edu/goldenPath/hg19/phastCons46way

SIFT http://sift.jcvi.org

PolyPhen-2 http://genetics.bwh.harvard.edu/pph2

MutationTaster www.mutationtaster.org

T1DBase http://www.t1dbase.org

BIODEF http://www.biopku.org/biodef

The Human Gene Mutation Database (HGMD) http://www.hgmd.org